

Memory system

Patent Number: ☐ DE4422786

Publication
date: 1995-01-12

Inventor(s): KUROSAWA HIROYUKI (JP); MASUZAKI HIDEFUMI (JP); INOUE YASUO (JP); ISONO SOICHI (JP); NINOMIYA TATUYA (JP); TAKAHASHI NAOYA (JP); HOSHINO MASAYUKI (JP); IWASAKI HIDEHIKO (JP)

Applicant(s): HITACHI LTD (JP)

Requested
Patent: ☐ JP7020994

Application
Number: DE19944422786 19940629

Priority Number
(s): JP19930162021 19930630

IPC
Classification: G06F13/12; G06F11/18; G06F12/08

EC
Classification: G06F11/20E, G06F11/20L4D, G06F11/16, G06F12/08B12

Equivalents: JP3264465B2


RECEIVED

SEP 23 2003

Technology Center 2100

Abstract

A memory system, which is connected to a mainframe, contains a plurality of first logic units (1), which are connected to a host device, a plurality of second logic units (2), which are connected to a memory device (5), a plurality of cache memories (3) and a common bus (4), which runs between these logic units and these memories and is functionally connected thereto. The first logic units, the second logic units and the cache memories all take the form of modules. The modules are releasably attached to the common bus which is arranged on a rear plane (71). The memory device can be formed by a plurality of memory units which are given small dimensions and are arranged in the form of a matrix. This means that the memory system is scalable. Since the first logic units, the second logic units and the cache memories are duplicated and the common bus is formed by two channels, the memory system can carry out reduced operation when specific components fail. Since the first logic units, the second logic units and the cache memories permit the components to be exchanged during operation, the memory system can be serviced without having to

halt its operation. 

Data supplied from the esp@cenet database - I2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-20994

(43) 公開日 平成7年(1995)1月24日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	3 0 1 B			
12/08	3 2 0	7608-5B		
13/12	3 3 0 T	8133-5B		

審査請求 未請求 請求項の数 9 O L (全 21 頁)

(21) 出願番号 特願平5-162021

(22) 出願日 平成5年(1993)6月30日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 二宮 龍也

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72) 発明者 増崎 秀文

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(72) 発明者 黒沢 弘幸

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74) 代理人 弁理士 武 顕次郎

最終頁に続く

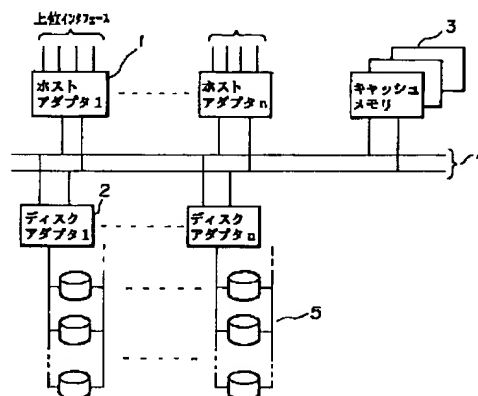
(54) 【発明の名称】 記憶システム

(57) 【要約】

【目的】 大形計算機の記憶システムで、システム規模を容易に拡張変更でき、システムの縮退及び活線挿抜による保守を可能とする。

【構成】 上位CPUと接続される複数のホストアダプタ（上位側インタフェース）1と、アレイドディスク5と接続される複数のディスクアダプタ（記憶装置側インタフェース）2と、これらのアダプタに共用される一時記憶用キャッシュメモリ3とは、これらアダプタ及びキャッシュメモリに共用されるコモンバス4上に挿抜自在に取り付けられる。規模を拡大するには、必要な数だけこれらアダプタ1、2及びキャッシュメモリ3を付加するだけでよい。アダプタ1、2、キャッシュメモリ及びコモンバスは二重化され、障害時の縮退運転を可能とし、また各アダプタ及びキャッシュメモリとコモンバスとの結合部は、活線挿抜可能としシステム無停止で保守点検部品交換を可能とする。

【図1】



(2)

特開平 7-20994

1

【特許請求の範囲】

【請求項 1】 上位装置に対するインタフェースを構成する複数の上位側接続論理装置と、記憶装置と、前記記憶装置に対するインタフェースを構成する複数の記憶装置側接続論理装置と、前記複数の上位側接続論理装置及び前記複数の記憶装置側接続論理装置間で転送されるデータを一時記憶するキャッシュメモリ装置とを有する記憶システムにおいて、前記複数の上位側接続論理装置、前記複数の記憶装置側接続論理装置、及び前記キャッシュメモリ装置は、これらの装置に共用されるコモンバスにより相互に接続されるように構成したことを特徴とする記憶システム。

【請求項 2】 前記複数の上位側接続論理装置、前記複数の記憶装置側接続論理装置、及び前記キャッシュメモリ装置は、いずれもモジュールで構成し、前記モジュールは、それぞれ、前記コモンバスに対し挿抜自在に取付けられるように構成したことを特徴とする請求項 1 記載の記憶システム。

【請求項 3】 前記コモンバスは、ブラッタ上に配設され、前記上位側接続論理装置、前記複数の記憶装置側接続論理装置、及び前記キャッシュメモリ装置を構成するモジュールの各々は、前記ブラッタに対し挿抜自在に取付けられるように構成したことを特徴とする請求項 1 または 2 記載の記憶システム。

【請求項 4】 前記上位側接続論理装置、前記記憶装置側接続論理装置、前記キャッシュメモリ装置、及び前記コモンバスは、いずれも少なくとも二重化されており、前記上位側接続論理装置、前記記憶装置側接続論理装置、前記キャッシュメモリ装置、及び前記コモンバスの一方により縮退運転が可能となるように構成したことを特徴とする請求項 1 ないし 3 のいずれか 1 記載の記憶システム。

【請求項 5】 前記二重化された上位側接続論理装置、記憶装置側接続論理装置、キャッシュメモリ装置は、いずれも活線挿抜ができるように構成したことを特徴とする請求項 4 記載の記憶システム。

【請求項 6】 前記記憶装置は、二重化された電源部を備えたことを特徴とする請求項 1 ないし 5 のいずれか 1 記載の記憶システム。

【請求項 7】 前記記憶装置は、複数の小形記憶装置を組み合わせたアレイ記憶装置で構成したことを特徴とする請求項 1 ないし 5 のいずれか 1 記載の記憶システム。

【請求項 8】 前記キャッシュメモリ装置は、キャッシュメモリ本体を持ち、前記コモンバスに直接取り付けられるキャッシュメモリモジュールと、キャッシュメモリを持つ増設用のキャッシュユニットとを有しており、前記キャッシュユニットは、前記コモンバスに直接挿抜自在に取り付けられる増設用のキャッシュポートパッケージを介して接続されるように構成したことを特徴とする請求項 1～7 のいずれか 1 記載の記憶システム。

2

【請求項 9】 前記上位側接続論理装置及び前記記憶装置側接続論理装置は、それぞれ、二重化されたマイクロプロセッサを有し、両マイクロプロセッサによりデータの比較チェックを行なうように構成したことを特徴とする請求項 1 ないし 8 のいずれか 1 記載の記憶システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、大形計算機システムやネットワークシステム等に接続される磁気ディスク装置、磁気テープ装置、半導体記憶装置、または光ディスク装置等の記憶装置を制御する記憶制御装置を含む記憶システムに係り、特に、システムの拡張性が高く縮退運転や活線挿抜対応の可能な記憶システムに関する。

【0002】

【従来の技術】従来、大形計算機に接続される記憶システムとして、例えば特開昭 61-43742 号公報に記載されているように、上位装置（CPU）に対するインタフェース（ホストアダプタ）、キャッシュメモリ、及び磁気ディスク装置等の記憶装置に対するインタフェース（ディスクアダプタ）の相互間をホットライン（専用線）で接続しているものが知られている。

【0003】図 20 は、従来の記憶システムの構成の概要を示す図である。同図において、201-1～201-n はそれぞれ複数の上位ホスト（CPU）に接続されるホストアダプタ（対上位論理モジュール）、202-1～202-n は、共有の大形ディスク装置 205 に接続されるディスクアダプタ（記憶媒体接続用論理モジュール）、203 は、複数のホストアダプタに共有のキャッシュメモリ、206 は同様に共有の管理メモリである。従来装置では、各ホストアダプタ 201-1～201-n とキャッシュメモリ 203 の間、キャッシュメモリ 203 と各ディスクアダプタ 202-1～202-n の間、各ホストアダプタ 201-1～201-n と管理メモリ 206 の間、並びに管理メモリ 206 と各ディスクアダプタ 201-2～201-n の間は、それぞれ別々のホットライン 207-1～207-n 及び 208-1～208-n によって接続されている。また、これらのホストアダプタ及びディスクアダプタの監視及び保守を行なう保守用プロセッサ（SVP、図示せず）も各々のホストアダプタ及びディスクアダプタにそれぞれ専用線を介して接続されている。

【0004】

【発明が解決しようとする課題】上記従来技術では、上位装置に対するホストアダプタ（対上位接続論理モジュール）と、記憶装置に対するディスクアダプタ（対記憶媒体接続論理モジュール）と、キャッシュメモリ（キャッシュメモリモジュール）との各間がホットラインで接続されているため、装置構成が複雑になると共に、ホストアダプタ、キャッシュメモリ、ディスクアダプタ、ディスク装置等、装置としての拡張性に乏しくいわゆるス

(3)

特開平7-20994

3

ケーラブル（拡張及び縮小自在）なシステム構成が得られなかった。また、システムを多重化することにより障害発生時等に縮退運転（2台のうち1台を停止し他の1台だけで運転するなど）や活線挿抜対応（システムを動作したままで基板や回路の部品等を押しかえるなど）を可能とすることがなにも配慮されておらず、このため、障害発生時の部品交換やシステムの制御プログラムをグレードアップするときには、システムを一時停止し対応しなければならない等の問題があった。

【0005】従って、本発明の目的は、上記従来技術の問題点を解決し、コモンバス方式を採用することにより、システム構成（規模）に応じてホストアダプタ、記憶装置アダプタ等の各論理モジュールやキャッシュメモリ及び記憶媒体を接続することでスケラブルなシステムを実現することができるようにすると共に、各論理モジュール、記憶媒体及びコモンバスの多重化により、縮退運転と各論理モジュール及び記憶媒体の活線挿抜対応とを可能とし、無停止で保守することができる記憶システムを提供することにある。

【0006】

【課題を解決するための手段】上記目的を達成するため、本発明は、上位装置に対するインタフェースを構成する複数の上位側接続論理装置と、記憶装置と、前記記憶装置に対するインタフェースを構成する複数の記憶装置側接続論理装置と、前記複数の上位側接続論理装置及び前記複数の記憶装置側接続論理装置間で転送されるデータを一時記憶するキャッシュメモリ装置とを有する記憶システムにおいて、前記複数の上位側接続論理装置、前記複数の記憶装置側接続論理装置、及び前記キャッシュメモリ装置は、これらの装置に共用されるコモンバスにより相互に接続されるように構成する。

【0007】前記複数の上位側接続論理装置、前記複数の記憶装置側接続論理装置、及び前記キャッシュメモリ装置は、いずれもモジュールで構成し、前記モジュールは、それぞれ、前記コモンバスに対し挿抜自在に取付けられるように構成する。

【0008】前記コモンバスは、ブラッタ上に配設され、前記上位側接続論理装置、前記複数の記憶装置側接続論理装置、及び前記キャッシュメモリ装置を構成するモジュールの各々は、前記ブラッタに対し挿抜自在に取付けられるように構成する。

【0009】前記上位側接続論理装置、前記記憶装置側接続論理装置、前記キャッシュメモリ装置、及び前記コモンバスは、いずれも少なくとも二重化されており、前記上位側接続論理装置、前記記憶装置側接続論理装置、前記キャッシュメモリ装置、及び前記コモンバスの一方により縮退運転が可能となるように構成する。

【0010】前記二重化された上位側接続論理装置、記憶装置側接続論理装置、キャッシュメモリ装置は、いずれも活線挿抜ができるように構成する。

4

【0011】前記記憶装置についても、同様に二重化された電源部を備えることができる。

【0012】前記記憶装置は、複数の小形記憶装置を組み合わせたアレイ記憶装置で構成することができる。

【0013】前記キャッシュメモリ装置は、キャッシュメモリ本体を持ち、前記コモンバスに直接取り付けられるキャッシュメモリモジュールと、キャッシュメモリを持つ増設用のキャッシュユニットとを有しており、前記キャッシュユニットは、前記コモンバスに直接挿抜自在に取り付けられる増設用のキャッシュポートパッケージを介して接続されるように構成することができる。

【0014】前記上位側接続論理装置及び前記記憶装置側接続論理装置は、それぞれ、二重化されたマイクロプロセッサを有し、両マイクロプロセッサによりデータの比較チェックを行なうように構成することができる。

【0015】なお、コモンバス上には、上記上位側接続論理モジュールとは別の形式の上位側インタフェースや、上記記憶装置側接続論理モジュールとは別の形式の記憶装置側インタフェースを置き換えたり増設したりすることもできる。

【0016】

【作用】上記構成に基づく作用を説明する。

【0017】本発明によれば、上位装置に対するインタフェースを構成する複数の上位側接続論理装置と、記憶装置と、前記記憶装置に対するインタフェースを構成する複数の記憶装置側接続論理装置と、これらの装置間で転送されるデータを一時記憶するキャッシュメモリ装置（複数の上位側接続論理装置及び複数の記憶装置側接続論理装置に共有されるキャッシュメモリ装置）とを有する記憶システムにおいて、前記複数の上位側接続論理装置、複数の記憶側接続論理装置、及びキャッシュメモリ装置は、これらの装置に共有されるコモンバスにより相互に接続されるように構成したので、上位側接続論理装置と記憶装置側接続論理装置とキャッシュメモリの増設または変更は、単にこれらをコモンバス上に追加しまたは変更して行くだけでよく、増設によるアップグレードが容易に達成できスケラブルなシステム構成を得ることができる。

【0018】また、これらの上位側接続論理装置、記憶装置側接続論理装置及びキャッシュメモリ装置は、モジュール化されて、コモンバスの配設されたブラッタに挿抜（着脱）自在に取り付けるようにしたので、これらの装置の必要な数量の増設作業も簡単である。

【0019】また、上位側接続論理装置、記憶装置側接続論理装置、キャッシュメモリ装置、及びこれらの間を接続するコモンバスは、二重化され、2系統に分けて配線されているので、これらの装置の一方に障害が発生したときでも、他方の装置を用いて縮退運転が可能である。なお、障害発生時に縮退運転状況を示す情報は、共有メモリに書き込まれる。

(4)

特開平7-20994

5

【0020】この場合、上位側接続論理装置、記憶装置側接続論理装置、及びキャッシュメモリ装置は、いずれも活線挿抜対応のコネクタ部を具備しているので、システムを停止することなく保守点検を行なって故障部品の交換を行なったり、増設用の部品を追加したりすることが可能である。

【0021】電源部も二重化され、それにより無停電電源装置を実現する。

【0022】記憶装置は、複数の小形記憶装置を組み合わせたアレイ形とされ、これにより従来の大形ディスク装置1台を用いたものに比べてアクセスタイムを短縮できる。

【0023】キャッシュメモリ装置は、コモンバスに直接取り付けられるキャッシュメモリモジュール（キャッシュメモリパッケージ）と、増設用のキャッシュユニットとで構成され、増設用のキャッシュユニットは、コモンバスに直接挿抜自在に取り付けられる増設用のキャッシュポートパッケージを介して必要数接続されるようになっているので、簡単に増減することができる。

【0024】異常により、高信頼性の記憶システムを得ることができる。

【0025】

【実施例】以下に、本発明の実施例を図面の図1から図18により説明する。

【0026】図1は本発明の概念図を示す。図1により、本実施例の概要を説明する。

【0027】1は、対上位CPU（ホスト）接続用論理モジュールであるホストアダプタ部、2は、対記憶媒体接続用論理モジュールであるディスクアダプタ部、3は、両モジュール間で転送されるデータを一時記憶するキャッシュメモリパッケージ（キャッシュメモリモジュール）、4はホストアダプタ1、ディスクアダプタ2、キャッシュメモリパッケージ3の間のデータ転送制御を司るコモンバス、5は、縦横にアレイ状に配置した記憶媒体である磁気ディスク群（以下「アレイディスク」という）である。ホストアダプタ1は、上位インタフェース側のデータ形式及びアドレス形式を記憶媒体インタフェース用のデータ形式及びアドレス形式に変換する手段と、これらを制御管理する二重化したマイクロプロセッサとを有している。ディスクアダプタ2は、記憶媒体へデータを格納するためのアドレス演算機能と、記憶データ保証用冗長データの生成機能と、記憶媒体構成情報を認識する機能と、これらを制御管理するマイクロプロセッサとを有している。

【0028】図1において、上位装置（CPU）から送られてきた書き込みデータは、ホストアダプタ1からコモンバス4を介して一度キャッシュメモリパッケージ3に書き込むことにより上位に終了報告を行い、その後の空き時間でキャッシュメモリパッケージ3からディスクアダプタ2を経由してアレイディスク5に書き込む。

6

【0029】また、上位装置からのデータ読み出し命令に対しては、キャッシュメモリパッケージ3上にデータが存在する場合はアレイディスク5からは読み出さず、キャッシュメモリパッケージ3上のデータを上位装置に転送する。一方キャッシュメモリパッケージ3上にデータが存在しない場合は、アレイディスク5からディスクアダプタ2によりコモンバス4を経由して一度キャッシュメモリパッケージ3に書き込まれた後同様にホストアダプタ1を経由して上位装置へ転送する。

【0030】コモンバス4上のホストアダプタ1、ディスクアダプタ2、キャッシュメモリパッケージ3各々はその接続数を任意に変えることができる。ホストアダプタ1の実装数を変えれば対上位接続バス数が増加し、上位ホストに対するデータ転送能力を高めることができる。ディスクアダプタ2の実装数を変えれば記憶媒体に対する接続バス数が増加し、記憶媒体に対するデータの書き込み／読み出しの転送能力を高めることができる。また、同時に記憶媒体の数も増加することができる。キャッシュメモリパッケージ3の実装数を変えればデータの一時格納場所であるキャッシュメモリの容量が増加し、記憶媒体の総容量に対するキャッシュメモリの容量の比率を高めることができるので、対上位装置からアクセスするデータがキャッシュメモリ上に存在する確率（以下「キャッシュヒット率」という）を高める等スケラブルな装置構成を実現できる。

【0031】図2は、図1の概念図の詳細な構成図を示したものである。図2は、図1の複数台のホストアダプタ及び複数台のディスクアダプタのうち、それぞれ1台だけを示し、他は図示を省略している。

【0032】ホストアダプタ1において、6はホストインタフェースの光信号を電気信号に変換する信号変換部、7は上位データフォーマットをアレイディスク5用フォーマットに変換するフォーマット変換部である。8はコモンバス4とのデータの授受を司るデータ転送制御部で、内部にバケット転送単位のデータを格納する記憶バッファを内蔵している。9は活線挿抜対応可能な小振幅電流駆動形バスドライバ（以下「BTL」という）である。

【0033】ホストからのデータ転送要求は10のマイクロプロセッサ（以下「MP」という）に引継がれ、ホストアダプタ1内のデータ転送制御は当MP10の管理下で行われる。

【0034】MP10はMP内の障害発生を検出するなど高信頼性を確保するために二重化されており、11のチェッカ部で同じ動作をする二重化されたMP10とMP10'を比較チェックしている。

【0035】12はMP10の制御プログラムを格納するブートデバイスで、このブートデバイス12には書き換え可能な大容量フラッシュメモリを採用しており、またMP10は必要に応じて13のローカルメモリに制御

(5)

特開平7-20994

7

プログラムをコピーして使用することにより、MP10のメモリアクセス時間の高速化を実現しており、図中破線で囲まれた部分29がチャネルアダプタモジュールであり、ホストアダプタ1には当モジュール29が2回路搭載してある。

【0036】ディスクアダプタ2において、14はアレイドディスクに書き込むデータをセクタ単位に格納するバッファメモリ、15はバッファメモリ14の制御及びデータ転送制御を行なうデータ制御バッファ部、16はアレイドディスク5に書き込むデータを保証するための冗長データを生成する冗長データ生成部、17はアレイドディスク5（ターゲット）に対するイニシエータ（SCSIのマスタ側インタフェース）である。

【0037】またディスクアダプタ2内のデータ転送制御は、ホストアダプタ1と同じ構成をとるMP周辺部（MP10、MP10'、チェック11、ブートデバイス12、ローカルメモリ13からなりディスクアダプタ用の制御プログラムを搭載する）の管理下で行なわれる。

【0038】アレイドディスク5は、図2では4つのディスク（ターゲット）しか示していないが、実際には1台のディスクアダプタ2に対し例えば4（横）×4（縦）～4（横）×7（縦）つのディスクで構成される。横列はECCグループ（Error Correction Group）を構成し、各ECCグループは例えば3つのデータディスクと1つのパリティディスクで構成される。更に、後述のように、このようなアレイドディスク5の1組に対し、二重化されたホストアダプタ二重化されたホストアダプタと二重化されたディスクアダプタを通じて、あるCPUからアクセスできるようになっている。そして、ホストアダプタの一方に障害が発生したときには、ホストアダプタの他方もしくはディスクアダプタの他方を通じて、同じCPUから同じアレイドディスクにアクセスすることができる。

【0039】キャッシュメモリパッケージ3において、18は各アダプタのMP10が共通にアクセス可能で種々の管理情報を記憶する共有メモリ部、19は共有メモリ制御部、20はキャッシュメモリ部、21はキャッシュメモリ制御部であり、両メモリ制御部19、21は共にメモリ書き込みデータ保証の為にECC生成回路、読み出しデータの検査及び訂正回路を内蔵し、キャッシュメモリパッケージ3全体で最大1GBのキャッシュ容量を実現しており、装置構成上は2面化して実装している。

【0040】キャッシュメモリ容量を更に増設する場合は、キャッシュメモリパッケージ3の代わりに（または、キャッシュメモリパッケージ3に加えて）22で示すキャッシュポートパッケージを実装し、23で示すブラッタ（基板差し込み板）間接続ケーブルを介して24で示すキャッシュユニットに接続し、（すなわち、増設

8

ユニット24内のキャッシュメモリには、キャッシュポートパッケージ22及びケーブル23を介してアクセスできるように構成され）、これによって、最大8GB2面までキャッシュ容量を増設することができる。図2では、キャッシュメモリパッケージ2を2面設けたのに加えて、キャッシュポートパッケージ22を実装し、これにケーブル24を介していくつかのキャッシュユニット24を接続した場合を示している。

【0041】以上述べたホストアダプタ1、ディスクアダプタ2、キャッシュメモリパッケージ3はコモンバス4を介してつながっているが、このコモンバス中、25は各アダプタのMP10が共有メモリをアクセスするためのマルチプロセッサバス（以下「Mバス」という）、26は高速データ転送を行う高速I/Oバス（以下「Fバス」という）である。

【0042】高速I/Oバス26は通常は64ビット幅で2系統同時に動作しているが、障害発生時はどちらか1系統のみでの縮退動作が可能であり、またMバス25に障害が発生した場合はFバス26のどちらか1系統を使用して動作可能である。

【0043】更に活線挿抜対応（挿抜の際、挿抜部品の負荷を小さくして挿抜を行なうことで、システムを稼働状態のまま挿抜を可能とする）のBTL9をコモンバス4のインターフェイスにすることで、ホストアダプタ1に障害が発生した場合、システムは自動的に本障害バスを閉塞し他のホストアダプタのバスを用いてアレイドディスク5に対し対上位（同じCPU）からのアクセスを継続する。保守員は、システム稼働状態において障害の発生したホストアダプタ1を取り除き、正常なホストアダプタ1をシステムに挿入し、27の保守用プロセッサ（以下「SVP」という）から28のLANを介して復旧の指示を与え、システムは交換されたホストアダプタ1の動作をチェックし正常であれば閉塞バスを復旧させることにより、無停止運転を実現している。なお、図中LANCは、LAN Controller（SVPインタフェースコントローラ）である。SVP27は、他のホストアダプタ及びディスクアダプタにも同様に接続され、監視及び保守が行なわれるようになっている。

【0044】また、各アダプタの制御プログラムに変更がある場合は、SVP27からLAN28を介してブートデバイス12内にある制御プログラムの内容を書き替えることにより無停止のアップグレードが可能である。

【0045】即ち、システムの制御プログラムをアップグレードを実施する場合は、まずホストアダプタ/ディスクアダプタの各モジュールを1モジュールずつ閉塞し、制御プログラムのアップグレードを行い再接続する。以上のように1モジュールずつの制御プログラムの入れ換え操作を繰り返すことにより、系全体の制御プログラム入れ換えが実施される。

【0046】図3は、図2に示した構成図に沿ってデー

(6)

特開平7-20994

9

タの流れとデータの保証を示した図である。

【0047】上位からアレイディスクにデータを書き込む場合、例えばESCON（光チャネルの商標名、IBM社）から、先ず書き込み先の記憶空間上の物理アドレス情報（以下「PA」という）が送られて来た後、データ（CKD（Count Key Data）フォーマット）+CRCコードが送られてくる。これらの光信号は信号変換部6で電気信号に変換すると共にパリティを生成し、フォーマット変換部7ではデータフォーマットをFBA（Fired Blocked Architecture）フォーマットに変換すると共にLRC（Longitudinal Redundancy Check、長手方向冗長度チェック）コードを付加し、更にPAをデータの一部として取り込んでアレイディスク上の論理アドレス（以下「LA」という）を生成した後これら総ての情報に対してパリティを付加してFバス26に送られる。

【0048】キャッシュパッケージ3では、Fバス26からのデータに対して誤り訂正可能なECCを付加してキャッシュメモリ20に書き込む。

【0049】ディスクアダプタ2では、Fバスからのデータに対して更にCRCコードが付加され、該データSCSIインターフェースを介してアレイディスク5に送られ、磁気ディスク装置個々にECCを付加して書き込みデータを保証している。

【0050】アレイディスク5からのデータ読み出しにおいても同様に、各チェックコードを元に読み出しデータの検査／訂正を行い信頼性を高めている。

【0051】以上のように、チェックコードはデータの長さ方向に対してはある長さ毎の水平チェック、データの垂直（幅）方向に対しては（例えばバイト単位の）垂直チェックで2重化されており、また転送が行われる領域間（図中一点鎖線）では当該2重化チェックコードのうち1つを必ずデータとして受け渡すことによりデータ保証に万全を期している。

【0052】図4は図1で述べたスケーラビリティを実現するための装置外観図であり、41はアレイディスクを制御する制御ユニット部、42はアレイディスクを実装するアレイユニット部で、本装置はこの2つのユニットで構成される。

【0053】図5は制御ユニット41の実装図で（a）は正面図、（b）は側面図を表わす。51はホストアダプタ1、ディスクアダプタ2、キャッシュメモリパッケージ3を実装する論理架部、52は停電時に揮発メモリであるキャッシュメモリ部に電源を供給するバッテリー部、53はキャッシュメモリ増設時にキャッシュユニット24及び増設メモリ用の追加バッテリーを実装するキャッシュメモリ増設部、54はSVP実装部、55は論理架に電源を供給する論理架用スイッチング電源、56はアレイディスクの構成（容量）が小規模の場合のアレイ

10

ディスク実装部、57はアレイディスク部に電源を供給するアレイディスク用スイッチング電源、58は両スイッチング電源55、57に電源を供給する商用電源制御部である。

【0054】図6は大容量アレイディスクを構成するときのアレイユニット部の実装図で（a）は正面図、（b）は側面図を表わす。

【0055】アレイディスク実装部56は、磁気ディスク装置を最大112台（8行×7列×2）実装可能であり、各磁気ディスク装置に障害が発生した場合の装置の入れ替えを容易にするために、装置の正面と背面の両面から挿抜可能となるような実装方式をとっている。

【0056】61はユニット全体の発熱を逃がすための冷却ファンで、冷却効果を高めると共に、騒音抑止の観点から小さな冷却ファンを使って小区分化し、床面より天井へ送風する構造をとっている。

【0057】図7は図5で説明した論理架部の接続方式図である。

【0058】71はコモンバス4をプリント配線したブラッタ（基板の挿し込み用の板）であり、72は各アダプタ、パッケージとブラッタ71を接続するためのコネクタである。

【0059】ホストアダプタ1、ディスクアダプタ2、キャッシュメモリパッケージ3の間のデータ転送はコモンバス4を介して行うため、各アダプタ、パッケージはコネクタ72上の任意のどの位置でも接続可能となり、ホストアダプタ1の実装数、ディスクアダプタ2の実装数を自由に変えることができる。

【0060】一方、キャッシュ容量を増設する場合はキャッシュメモリパッケージ3をキャッシュポートパッケージ22に変えて実装するか、または図7に示すように、キャッシュメモリパッケージ3に加えてキャッシュポートパッケージ21を実装し、これに、接続ケーブル23を介してキャッシュユニット43（図2の24に相当）に接続することにより、もとの2GBの容量に加えて更に最大8GB2面分のキャッシュメモリ容量を拡張できる。

【0061】図8は図5で示した論理架部の実装イメージ図である。

【0062】図8で、コモンバス4は、ブラッタ71上を左右方向にプリント配線されており、このブラッタ71に対して、キャッシュポートパッケージ22の基板（CP）の取付部、キャッシュメモリパッケージ3の基板（C）の取付部、ホストアダプタモジュールの基板（H）の取付部、及びディスクアダプタモジュールの基板（D）の取付部が設けられ、図の矢印84で示すように、各基板は、挿抜操作面側から着脱されるようになっていて、ブラッタ71に差し込まれるとコモンバス4と電気接続されるものである。

【0063】81は、ホストアダプタ1の基板上の下方

(7)

特開平7-20994

11

部に実装されて、対上位インターフェイスを司る光コネクタ部、82はディスクアダプタ2の基板上の下方部に実装されて、アレイディスク5と接続するSCSIコネクタ部、83はキャッシュポートパッケージ22を実装したときの接続ケーブル23用の接続コネクタ部である。85は、キャッシュメモリパッケージ3の基板(C)の下方部に取付けたキャッシュメモリ本体(図2のキャッシュメモリ20)である。

【0064】各コネクタ部は、障害発生等で各アダプタ、パッケージを挿抜する際の操作性を向上させるため、接続コネクタ83を除き、操作面84側へは実装せず、ブラッタ71の接続側に集中実装している。

【0065】図9は本発明のソフトウェア構成を示した図である。

【0066】91はホストアダプタ1のブートデバイス12に書き込まれるチャンネルアダプタ制御プログラム(以下「CHP」という)、である。また、ディスクアダプタ2のブートデバイス12に書き込まれるディスクアダプタ制御プログラムのうち、92はアレイディスク固有の処理およびキャッシュメモリとアレイディスク間のデータ転送制御を受け持つディスクアダプタマスタ制御プログラム(以下「DMP」という)、93はDMP92の制御管理下でキャッシュメモリ20とアレイディスク5の間のデータ転送制御を受け持つディスクアダプタスレーブ制御プログラム(以下「DSP」という)である。

【0067】ディスクアダプタ2のブートデバイス12には、DMP92とDSP93の2種類が書き込まれているが、装置構成上nセットのディスクアダプタでアレイディスクにアクセスする場合、そのうちの2セットがDMP92として動作(2重化)し、残るn-2のディスクアダプタがDSP93として動作する。

【0068】94はSVP27に搭載するSVP制御プログラムで、CHP91、DMP92、DSP93を監視及び保守するとともに、各制御プログラムの更新時はSVP27から更新したいMPの制御プログラムを直接、または他のMPから当該MPの制御プログラムを更新することができる。

【0069】図10はデータの流れに基づいた図9で示したソフトウェア構成の機能分担を示した図である。

【0070】CHP91は、上位からのアドレス形式及びデータ形式を下位アドレス形式及びデータ形式に変換し、キャッシュメモリに書き込む。101はセグメント、102はブロック、103はアレイディスク5上の磁気ディスク1台当りに書き込むデータ量を表すストライプである。DMP92は、キャッシュメモリ上からストライプ単位にデータを読み出し、下位アドレスをアレイディスクの行NO、列NO、FBA、ブロック数に変換し、DSP93でアレイディスクにデータを書き込む。

12

【0071】また、DMP92はアレイディスク5の構成情報も管理している。

【0072】以上のように、各制御プログラムを機能分担することにより、上位インタフェースをSCSIやファイバーチャネル等に変更する場合はCHP91のみ、またアレイディスク構成を変更(ディスクの行数/列数、RAID(Redundant Array Inexpensive Disk)方式等)する場合はDMP92のみの変更で対応可能であり、ホストアダプタ1、ディスクアダプタ2の接続変更に合わせて各制御プログラムを書き替えることで、スケラビリティを実現するとともに、ソフトウェア開発の負荷も軽減している。

【0073】図11はコモンバス4の2重化の考え方と縮退動作を説明した図である。

【0074】111はコモンバス4の使用権を獲得することのできるバスマスタ(MP10を搭載しているホストアダプタ1又はディスクアダプタ2)、112はバスマスタ111からのアクセス要求を受けるバススレーブ(キャッシュメモリパッケージ)である。

【0075】Fバス26は通常動作状態では64ビットバス(200MB/S)2系統を同時に動作させ400MB/Sを実現しており、各バス系統はパリティチェック又はタイムアウトで障害を検出可能である。障害発生時はバスマスタ111は各自縮退状態に入り、残る1系統を使ってバススレーブをアクセスすると共に、この時の縮退情報は共有メモリ18上の管理エリアに登録される。

【0076】またコモンバス内のシステム制御信号(バスリセット等)は信号線を3重化しており、通常動作時は3線一致、縮退動作時は2線一致(多数決)方式を採用することにより信頼性を高めている。

【0077】図12は装置各部位における多重化と縮退運転を示した図である。

【0078】121は2ポート化されたチャンネルバスであり、ホストアダプタ1にはチャンネルアダプタ29が2モジュール、対上位用のチャンネルバスが4バス実装しており、障害発生時は交替チャンネルアダプタ(CHP)、交替チャンネルバスを使用して縮退運転に入る。

【0079】122はディスクアダプタ2とアレイディスク5の間のインタフェースを司るSCSIバスで、1行の磁気ディスク群に対して別のディスクアダプタ2からもアクセス可能のように2重化しており、当バスに障害が発生した場合は交替SCSIバスを使用して縮退運転に入る。また、アレイディスクマスタ制御を行うDMP92も2重化しており、障害発生時は交替DMP92を使用して縮退運転に入る。

【0080】共有メモリ18、キャッシュメモリ20も2重化しており、共有メモリに障害が発生した場合は残るもう一方の使用して縮退運転に入り、キャッシュメモ

13

りに障害が発生した場合はライトベンディングデータ（キャッシュメモリ上に残っているデータ）をディスクにデステージし障害発生メモリ部位を除いたメモリで縮退運転を行う。

【0081】アレイディスク5上の磁気ディスクに障害が発生した場合は、当該磁気ディスクを切り離し予備の磁気ディスクに修復しながら読み出し書き込み動作を行う。

【0082】図13は装置の電源系の多重化と縮退運転を示した図である。

【0083】商用電源制御部58は各々独立したAC入力で2重化して、論理架用スイッチング電源55及びアレイディスク用スイッチング電源57にそれぞれ供給しているため、障害発生時はもう片方の商用電源制御部58で縮退運転に入る。

【0084】131は上位ホストからの電源ON/OFFの遠隔制御や商用電源制御部58、両スイッチング電源等の電源回路を制御する電源制御回路（以下「PCI」という）である。

【0085】論理架用スイッチング電源55は冗長運用として必要数より2回路多く実装し電源共通バスを介して論理架51及びバッテリー52に供給することにより、当スイッチング電源55が2回路故障しても動作可能である。

【0086】同様に列単位の磁気ディスク群に供給するにアレイディスク用スイッチング電源57も、冗長運用として2回路多く実装し電源共通バスを介して供給することにより、当スイッチング電源57が2回路故障しても動作可能であり、さらに両スイッチング電源55、57を2重化するよりも安価な構成に仕上げる事ができる。

【0087】また停電時においては、2重化されたバッテリー52から電源共通バスを介して論理架内の揮発メモリであるキャッシュメモリ及びPCI131に供給され、片方のバッテリーが故障しても動作可能である。

【0088】図14及び図15はアレイディスクに使用する磁気ディスク装置単体の記憶容量別にアレイディスクを構成したときのシステム性能を比較した図である。

【0089】図14はそれぞれ異なる磁気ディスク装置を使用して同一容量のアレイディスクを実現した場合の構成を示しており、項番141が3GBの磁気ディスク装置（3.5インチ径のディスクを使用）、項番142が4.0GBの磁気ディスク装置（5インチ径のディスクを使用）、項番143が8.4GBの磁気ディスク装置（6.4インチ径のディスクを使用）を使用している。アレイ構成は、ディスク装置141が14枚のデータディスクの2枚のパリティディスク、ディスク装置142が14枚のデータディスクと4枚のパリティディスク、ディスク装置143が14枚のデータディスクと2枚のパリティディスクで構成した場合である。

(8)

特開平7-20994

14

【0090】図15は各磁気ディスク装置141、142、143についての毎秒当りのI/O命令発行件数と平均応答時間の関係を示しており、アレイディスクシステムとしてのトランザクション性能を向上させるためには、小容量（小径）の磁気ディスク装置を使用してアレイ構成を大きくすることが最も性能を引き出せることから、本発明に於ては3.5インチ磁気ディスク装置141を採用してアレイディスクシステムを実現している。従って、同じ記憶容量の磁気ディスク装置を、従来のように大形磁気ディスク装置1台で構成するのと、複数台の小形磁気ディスク装置のアレイで構成するのとでは、後者の小形磁気ディスク装置を多数用いたアレイ構成のものの方が、平均アクセスタイムを短縮できる点で有利である。

【0091】以上説明してきたスケーラブルなアーキテクチャを使用して実現できる装置モデル構成例を図16～図19に示す。

【0092】図16は、共通バス4上のディスクアダプタ2の実装数を減らし、更にキャッシュポートバッファ22を実装し、接続ケーブル23を介してキャッシュユニット24に接続することにより、キャッシュヒット率の高める高性能大容量キャッシュメモリ付小形ディスクアレイを実現した時の構成図である。

【0093】またディスクアダプタ2を実装しないで、ホストアダプタ1とキャッシュメモリのみで構成した場合（図中の破線内の構成）は、記憶媒体が磁気ディスクから半導体メモリに代わり、更に高速データ転送可能な高性能の半導体ディスク装置を実現する。

【0094】図17はディスクアダプタ2を最大構成とし、キャッシュポート3を実装し又はキャッシュポート22を実装し接続ケーブル23を介してキャッシュユニットを接続することにより、高性能大容量キャッシュメモリ付大形ディスクアレイを実現した時の構成図である。

【0095】図18はホストアダプタ1の対上位インターフェースをSCSI/ファイバチャネル等のインターフェースに変えて、ディスクアダプタ2の実装数を減らし、更にFバス26のビット幅を半分に縮小した2系統で構成することにより、オープン市場をターゲットにした無停止運転の高性能フォールトトレラント（高信頼性）サーバシステムを実現した時の構成図である。

【0096】図19は図18の構成を元に2重化、活線挿抜を考慮せずに、最もシンプルな構成をとることによって安価なオープン市場向けのサーバシステムを実現した時の構成図である。なお、図中、4D+1Pは、データディスク4枚とパリティディスク1枚の趣旨である。

【0097】以上の実施例において、共通バス4上に、更に光ディスクアダプタ（光ディスク用接続論理モジュール）を介して光ディスク装置を接続し、磁気テープ制御装置（磁気ディスク接続論理モジュール）を介し

(9)

特開平7-20994

15

て磁気テープ装置を接続し、あるいは半導体記憶装置接続論理モジュールを介して半導体記憶装置を接続することができる。また、コモンバス4上に別の形式のホストアダプタを介してワークステーションを接続することもできる。このように、コモンバス上に、種々の形式の記憶装置に対する記憶媒体アダプタを接続することができる。

【0098】

【発明の効果】以上詳しく説明したように、本発明によれば、上位装置に対するインタフェースを構成する複数の上位側接続論理装置と、記憶装置と、前記記憶装置に対するインタフェースを構成する複数の記憶装置側接続論理装置と、これらの装置間で転送されるデータを一時記憶するキャッシュメモリ装置（複数の上位側接続論理装置及び複数の記憶装置側接続論理装置に共有されるキャッシュメモリ装置）とを有する記憶システムにおいて、前記複数の上位装置側接続論理装置、複数の記憶装置側接続論理装置、及びキャッシュメモリ装置は、これらの装置に共有されるコモンバスにより相互に接続されるように構成したので、上位側接続論理装置と記憶装置側接続論理装置とキャッシュメモリの増設または変更は、単にコモンバス上にこれらの装置等を追加または変更して行くだけでよく、増設によるアップグレードが容易に達成できスケラブルなシステム構成を得ることができる。また、これらの上位側接続論理装置、記憶装置側接続論理装置及びキャッシュメモリ装置は、モジュール化されて、コモンバスの配設されたブラッタに挿抜（着脱）自在に取り付けるようにしたので、これらの装置の必要な数量の増設作業も簡単であるという効果がある。

【0099】また、上位側接続論理装置、記憶装置側接続論理装置、キャッシュメモリ装置、及びこれらの間を接続するコモンバスは、二重化され、2系統に分けて配線されているので、これらの装置の一方に障害が発生したときでも、他方の装置を用いて縮退運転が可能である。この場合、上位側接続論理装置、記憶装置側接続論理装置、及びキャッシュメモリ装置は、いずれも活線挿抜対応のコネクタ部を具備しているため、システムを停止することなく保守点検を行なって故障部品の交換を行ったり、増設用の部品を追加したりすることが可能であるという効果がある。

【0100】更に、記憶装置は、複数の小形記憶装置を組み合わせたアレイ形とされ、これにより従来の大形ディスク装置1台を用いたものに比べてアクセスタイムを短縮できるという効果がある。

【0101】また、キャッシュメモリ装置は、コモンバ

16

スに直接取り付けられるキャッシュメモリモジュール（キャッシュメモリパッケージ）と、増設用のキャッシュユニットとで構成され、増設用のキャッシュユニットは、コモンバスに直接挿抜自在に取り付けられる増設用のキャッシュポートパッケージを介して必要数接続されるようになっているので、簡単に増減することができるという効果も得られる。

【0102】以上により、高信頼性の記憶システムを得ることができる。

10 【図面の簡単な説明】

【図1】本発明の実施例の概要を示す概念図である。

【図2】本発明の一実施例の記憶システムの詳細な構成図である。

【図3】図2の構成図に沿ったデータの流れとデータ形式を示した図である。

【図4】本発明の一実施例の装置外観図である。

【図5】本発明の一実施例の装置における制御ユニット部の実装方式図である。

20 【図6】本発明の一実施例の装置におけるアレイディスクユニット部の実装方式図である。

【図7】本発明の一実施例の装置における論理架部の接続方式図である。

【図8】本発明の一実施例の装置における論理架部の実装方式図である。

【図9】本発明の実施例に適用されるソフトウェア構成図である。

【図10】本発明の実施例によるデータの流れとソフトウェアの機能分担を示した図である。

30 【図11】本発明の実施例によるコモンバスの2重化と縮退動作を示した図である。

【図12】本発明の実施例による装置各部位の2重化と縮退運転を示した図である。

【図13】本発明の実施例による装置の電源系の多重化と縮退運転を示した図である。

【図14】アレイディスクに使用する磁気ディスク装置単体のディスク構成を示す図である。

【図15】磁気ディスク装置の記憶容量とアレイディスクのシステム性能を示した図である。

40 【図16】高性能大容量キャッシュメモリ付小形ディスクアレイの構成図である。

【図17】高性能大容量キャッシュメモリ付大形ディスクアレイの構成図である。

【図18】高性能フォールトトレラントサーバシステムの構成図である。

【図19】低価格サーバシステムの構成図である。

【図20】従来の記憶システムの概略構成図である。

(10)

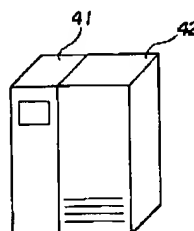
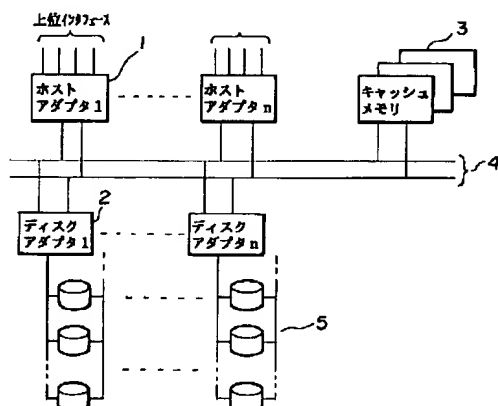
特開平7-20994

【图 1】

【図4】

【图 1】

【图4】

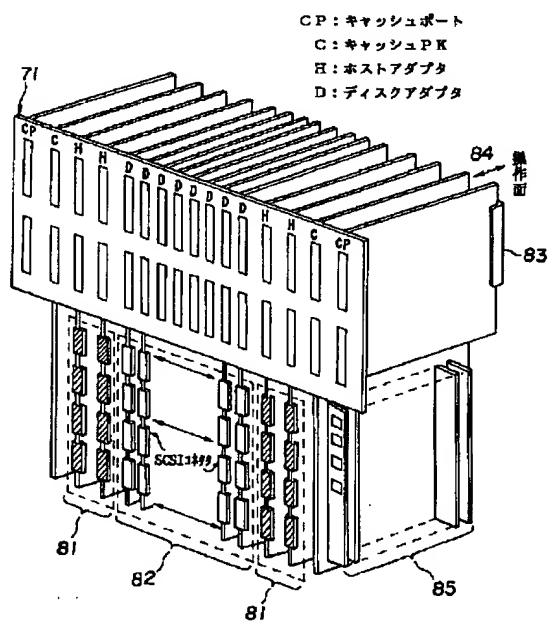
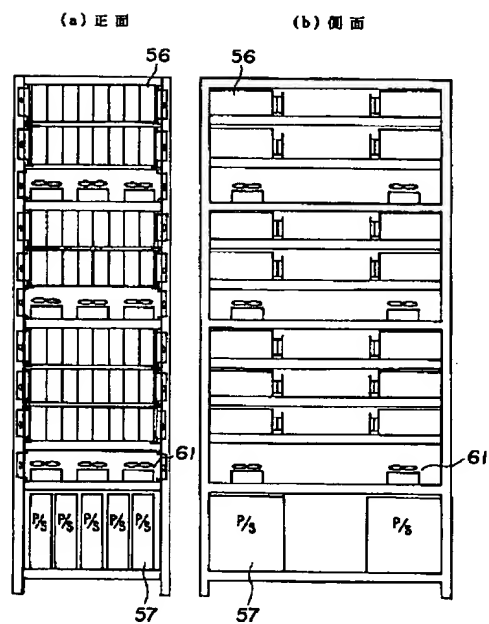


【図6】

【图8】

【図 6】

【图8】

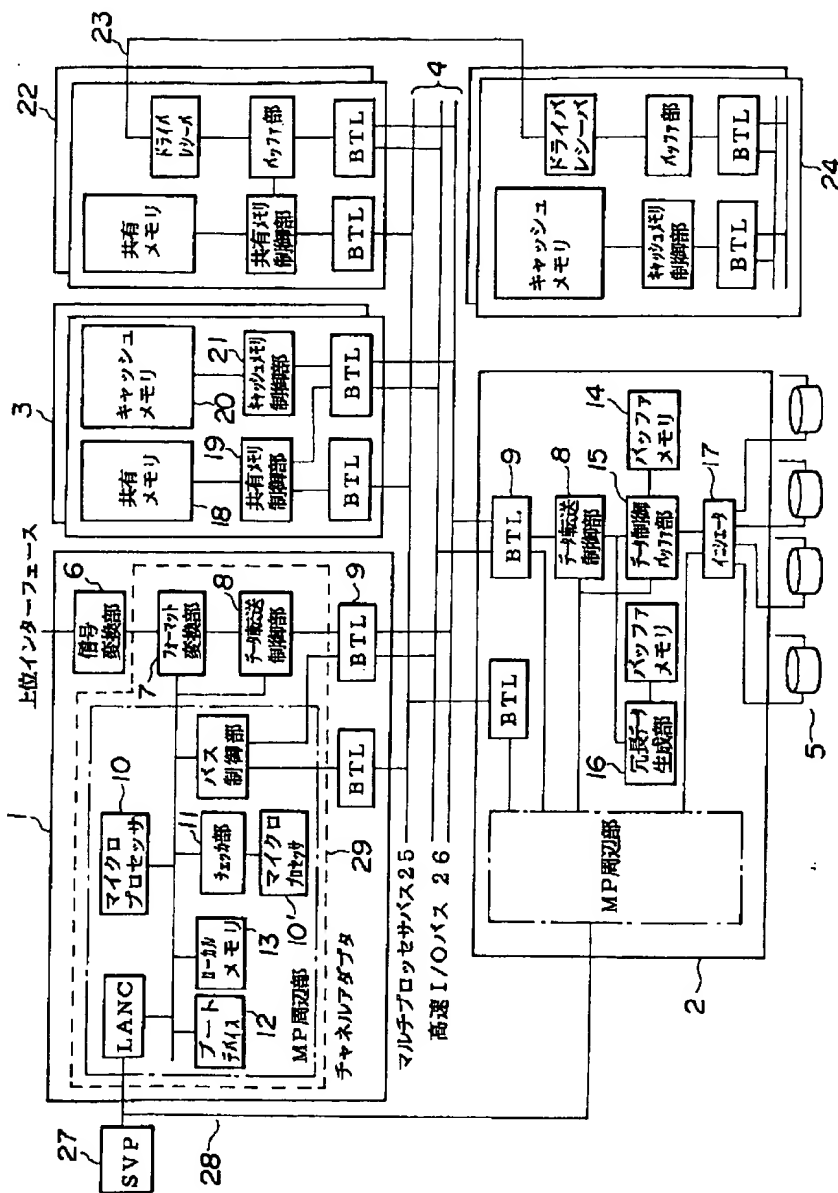


(11)

特開平7-20994

【図2】

【図2】

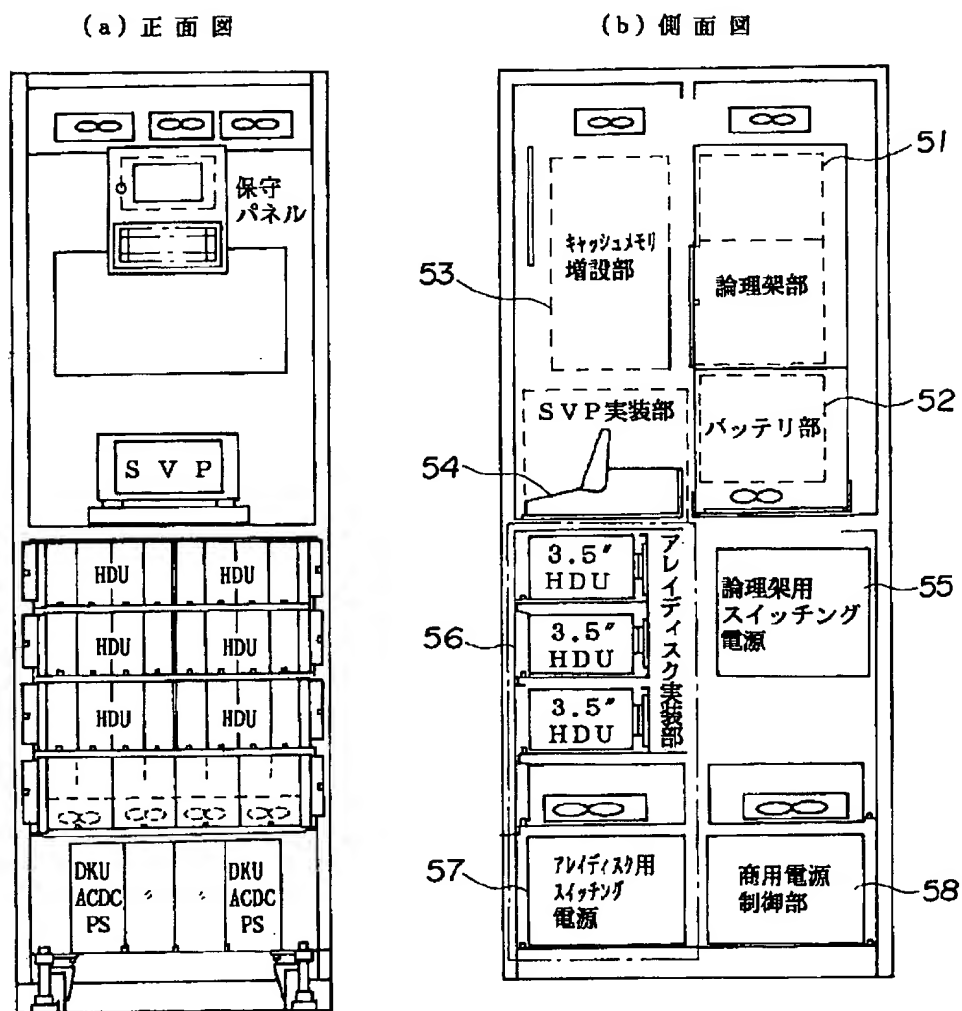


(13)

特開平7-20994

【図5】

【図5】



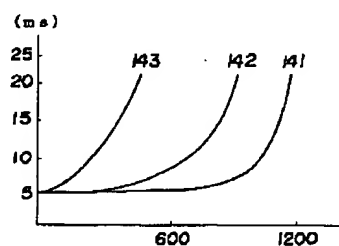
【図14】

【図15】

【図14】

【図15】

項番	磁気ディスク単体容量	フレイ構成	フレイ容量
141	3.0GB (3.5インチ)	(14D+2) × 5	約220GB
142	4.0GB (5インチ)	(14D+2) × 4	
143	8.4GB (6.4インチ)	(14D+2) × 2	

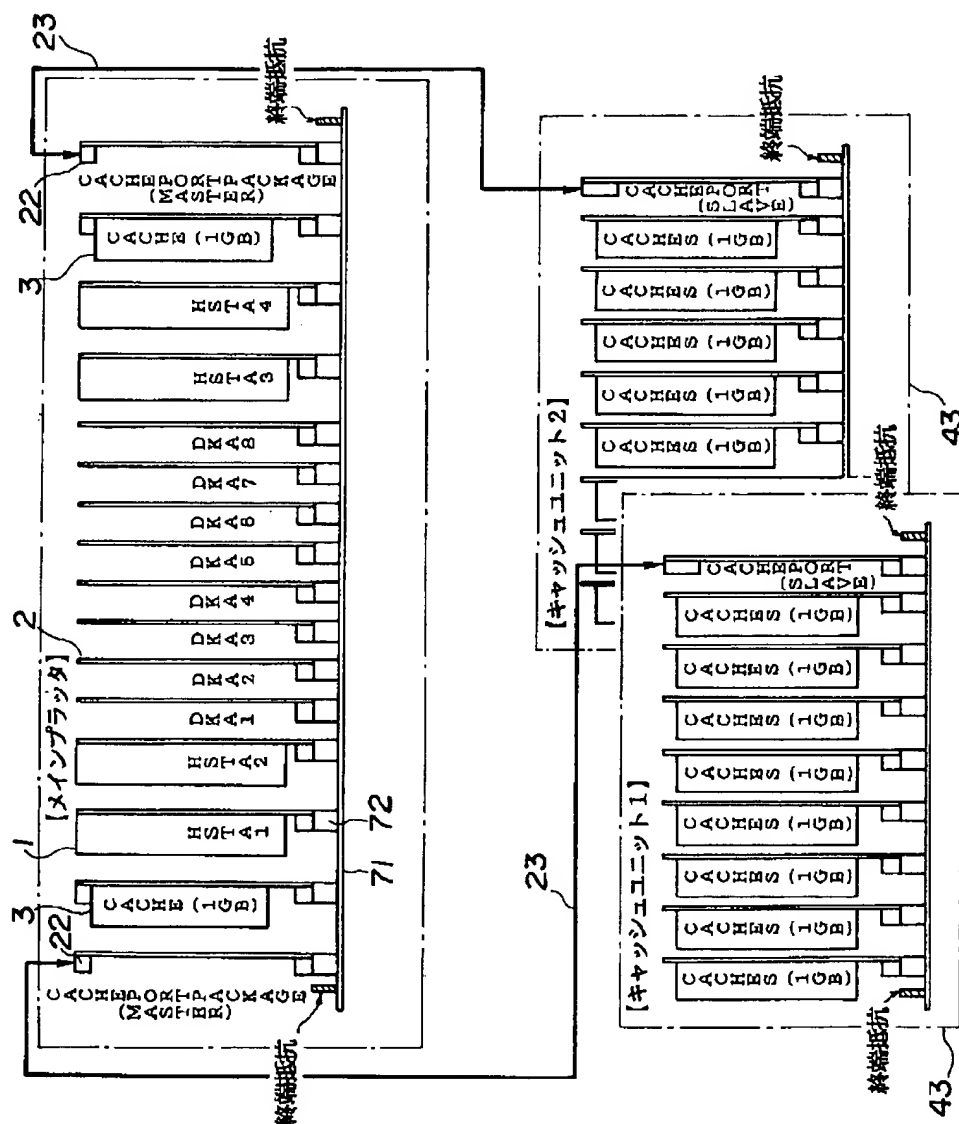


(14)

特開平7-20994

【図7】

【図7】

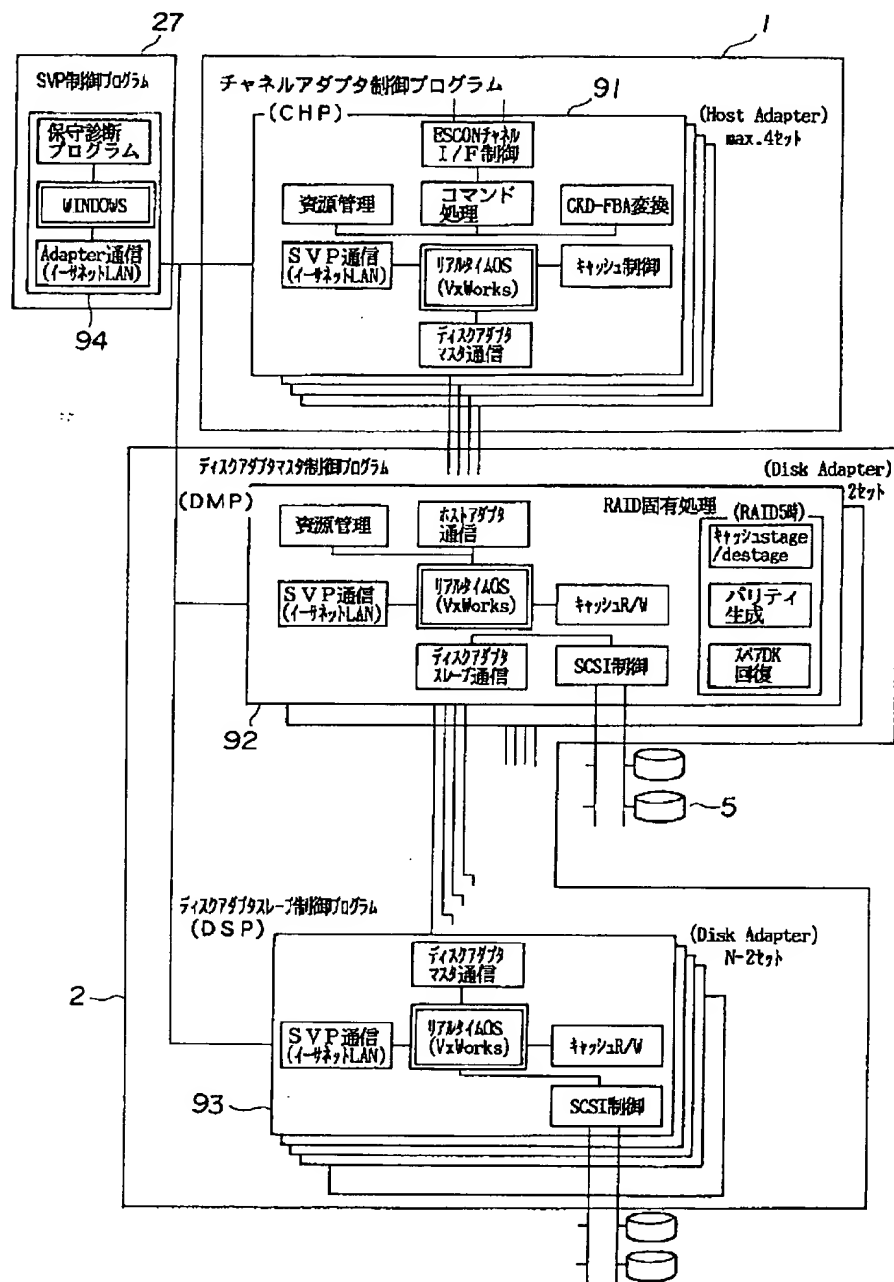


(15)

特開平7-20994

【図9】

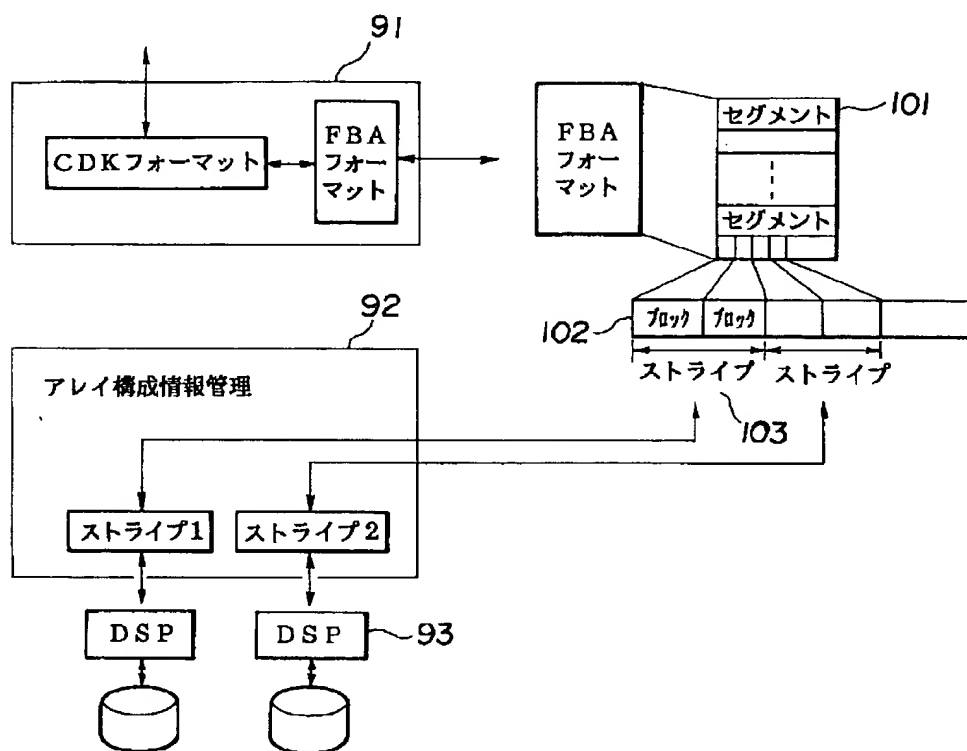
【図9】



(16)

特開平7-20994

【図10】



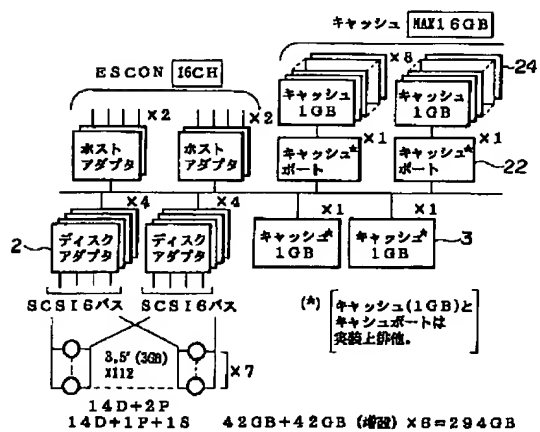
【図10】

【図17】

【図18】

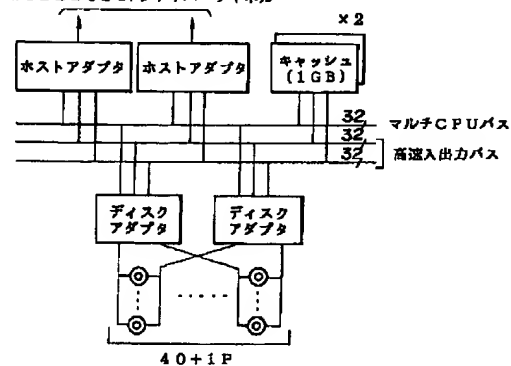
【図17】

【図18】



- ・RAID5/3/1/0
- ・システムキャッシュサポート
- ・フォールトトレラント

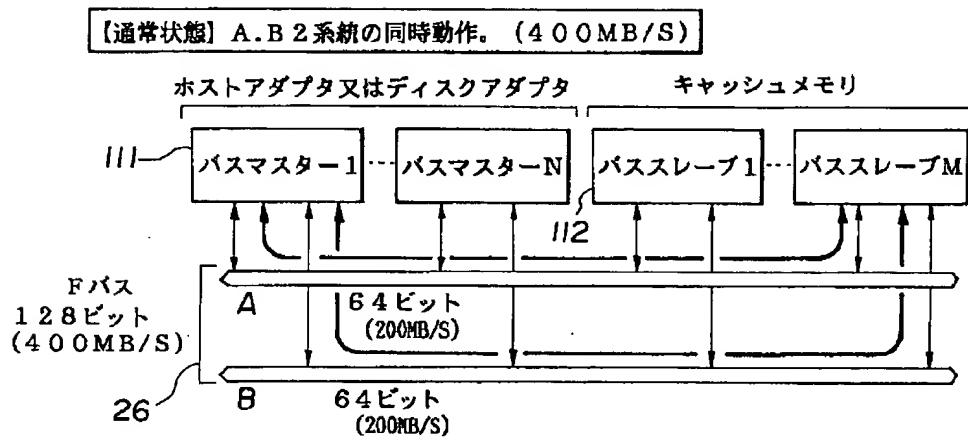
WIDE SCSI/ファイバーチャネル



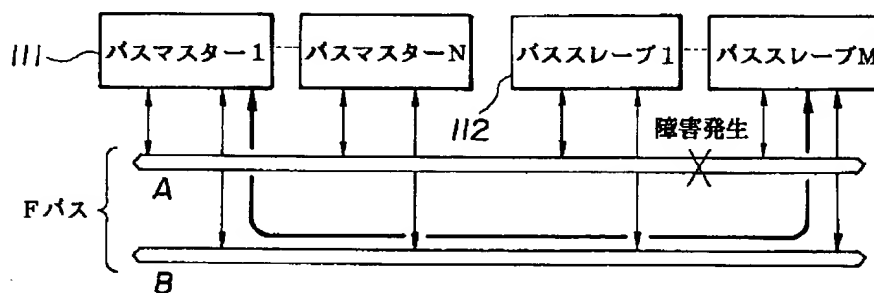
(17)

特開平7-20994

【図11】

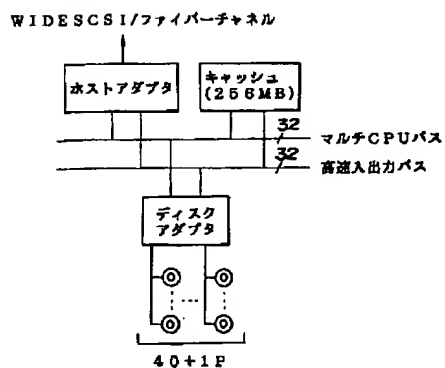


【縮退状態】 A.B 1系統にて縮退動作。(200MB/S転送)



【図19】

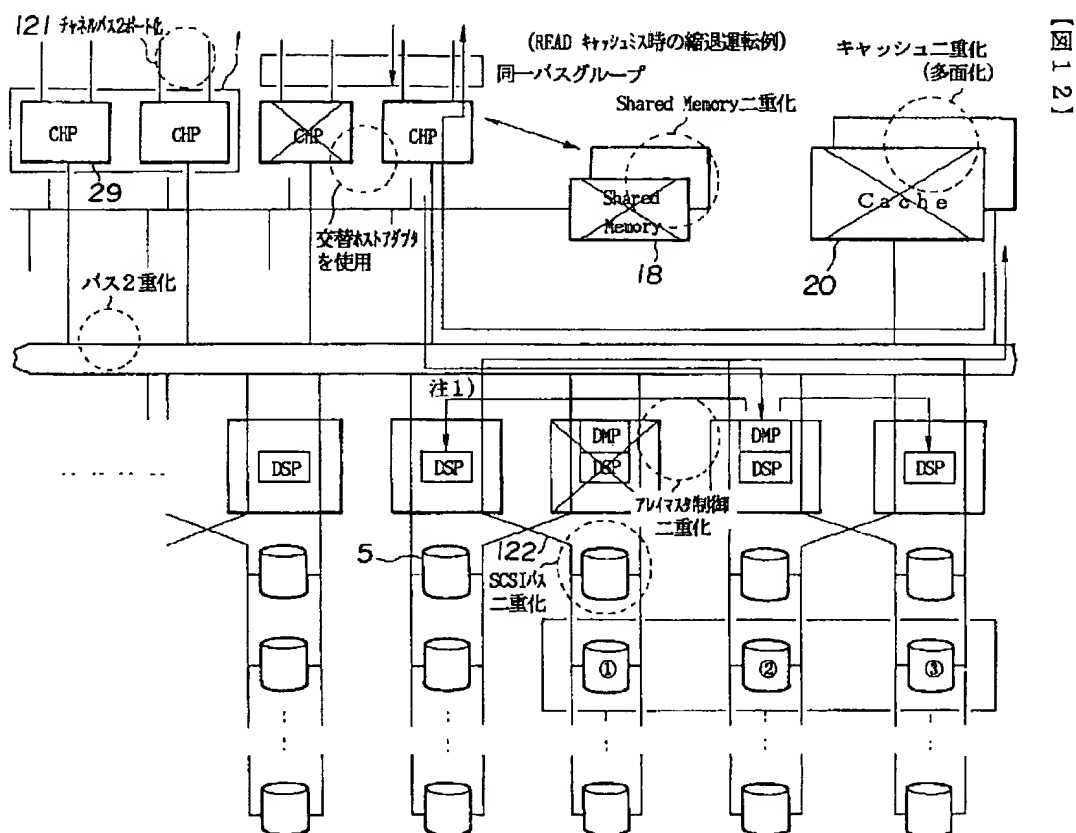
【図18】



(18)

特開平7-20994

【図12】

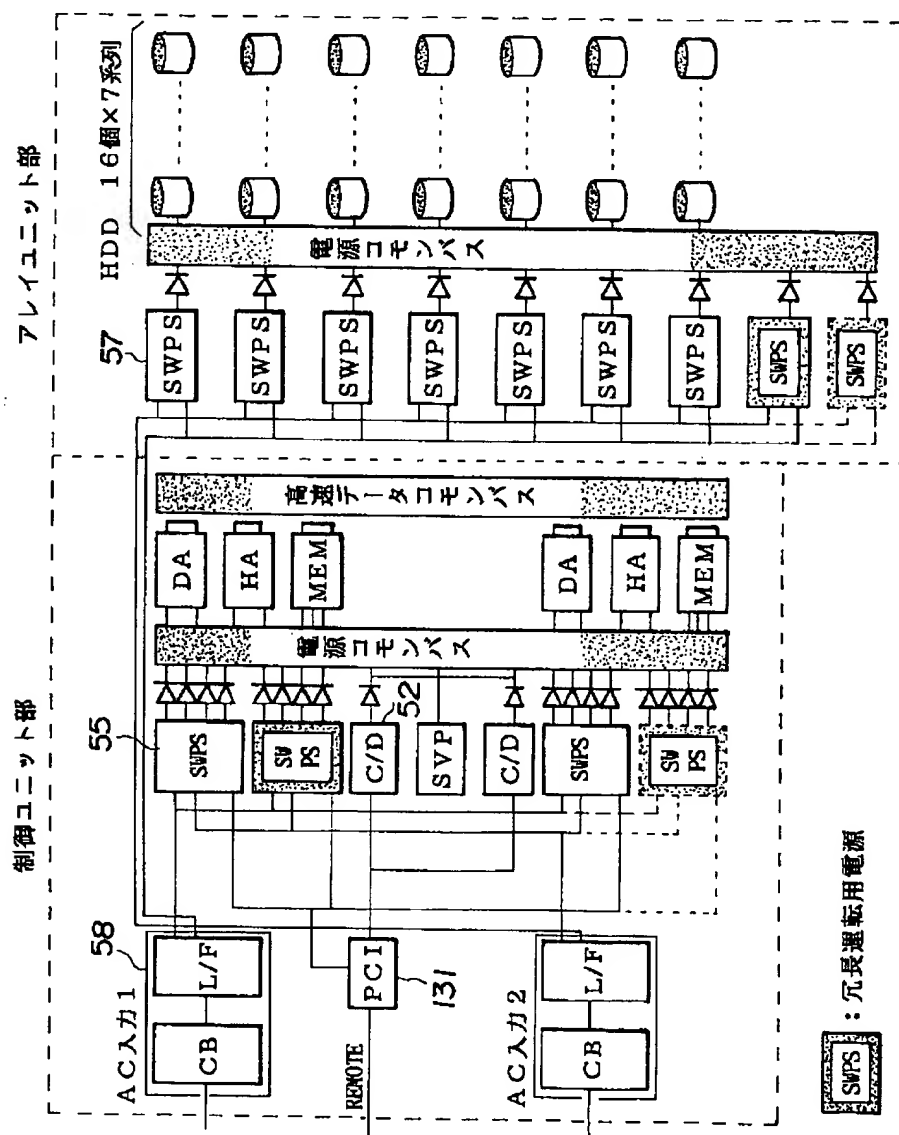


(19)

特開平7-20994

【図13】

【図13】



特開平7-20994

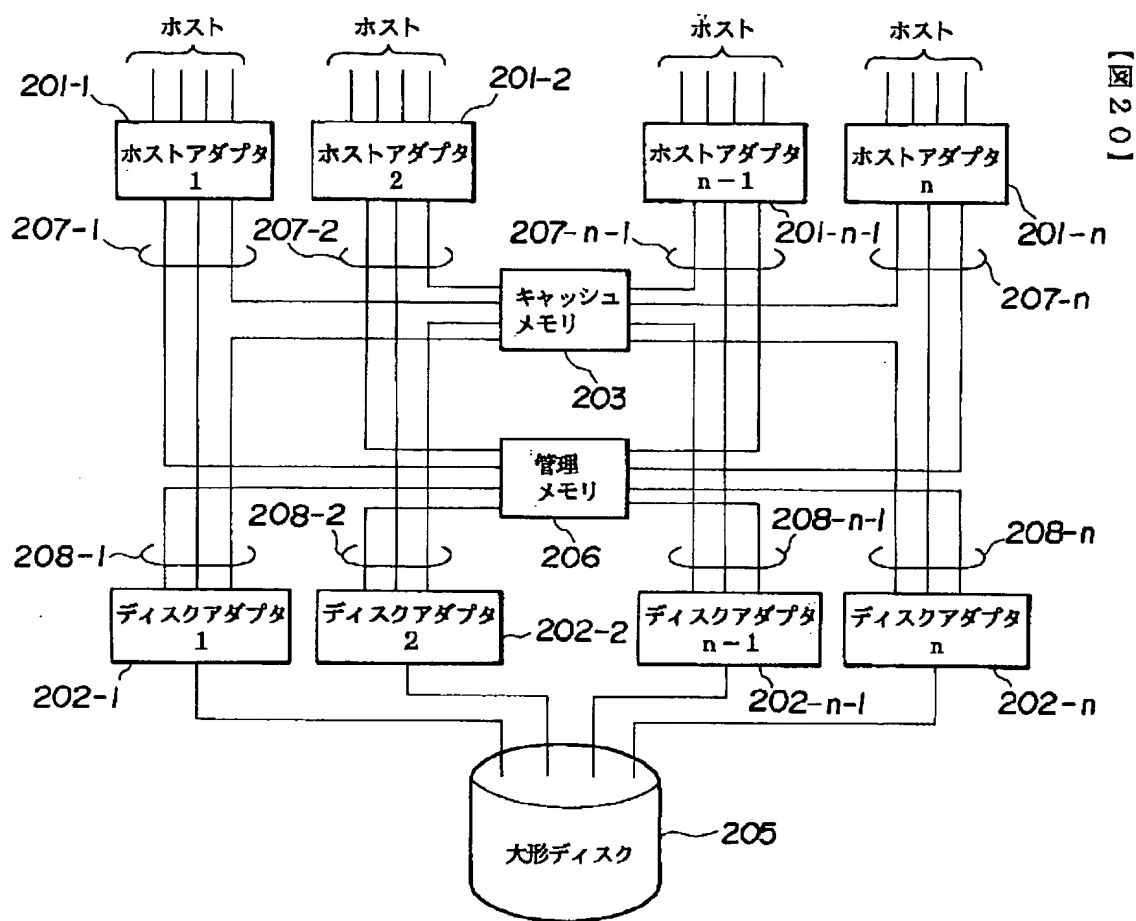
【図 16】



(21)

特開平7-20994

【図20】



フロントページの続き

(72)発明者 高橋 直也
 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内

(72)発明者 井上 靖雄
 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内

(72)発明者 岩崎 秀彦
 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内

(72)発明者 星野 政行
 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内

(72)発明者 磯野 聡一
 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内